CS 7545: Machine Learning Theory	**Fall 2021**
Lecture 9	
Name: Tyler LaBonte	**Professor:** Santosh S. Vempala

1 Spectral Clustering II

We've analyzed one spectral clustering algorithm using SVD, and while this algorithm is generally pretty good, it still relies on distance clustering in the SVD subspace. This means it cannot handle certain cases such as concentric circles. Now, we will see a spectral clustering algorithm that is able to handle these cases, with close relation to Markov chains.

In a general setting, our goal is to cluster the vertices of a graph $G = (V, E)$ with weights on the edges. The adjacency matrix $\mathbf{A}$ is symmetric, and a_{ij} is a non-negative similarity measure between $i, j \in V$. We would like "more similar" edges to be included in the same cluster. One idea is to use the minimum cut in the graph, but this does not exactly capture the idea of clustering (say the graph is two large clusters, then the minimum cut may be within one of these clusters). Instead, we will normalize the cut by the size of the smallest component, called the *conductance*.

Definition 1.1: Conductance

For $S \subseteq V$, let $a(S) = \sum_{i \in S, j \in V} a_{ij}$ be the sum of the degrees of the vertices in S. Then

$$\phi(S) = \frac{\sum_{i \in S, j \in \overline{S}} a_{ij}}{\min(a(S), a(\overline{S}))}. \tag{1}$$

Notice $\phi(S) \in [0, 1]$. We could replace the min in the denominator with the product up to a small constant, but it doesn't really matter. Then,

$$\phi(G) = \min_{S \subseteq V} \phi(S). \tag{2}$$

It will be useful to have the rows of the adjacency matrix sum to one, so we normalize $\mathbf{A} \to \mathbf{B}$ with $b_{ij} = a_{ij}/a(i)$. Note $\mathbf{B} \geq 0$ but it is non-symmetric. $\mathbf{B}$ has a few important properties that we will use:

1. $\mathbf{B}$ has largest eigenvalue 1 (proof by triangle inequality and rows sum to 1).
2. The left eigenvector is proportional to the degrees: $\mathbf{B}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$ with $\boldsymbol{\pi}(i) \propto \boldsymbol{a}(i)$.
3. $\pi_i b_{ij} = \pi_j b_{ji}$.

These properties may look familiar – it means that $\boldsymbol{\pi}$ is the stationary distribution for the time-reversible Markov chain with transition matrix $\mathbf{B}^\top$. The conductance has a nice interpretation here: it is the conditional escape probability of the Markov chain from the set $S \subseteq V$.

Theorem 1.1: Spectral Conductance

Let $\mathbf{B}^\top$ be an $n \times n$ transition matrix for a time-reversible Markov chain. Let $\boldsymbol{v}$ be the second eigenvector of $\mathbf{B}$ with eigenvalue $\lambda_2 \leq 1$ and components $v_{i_1} \geq v_{i_2} \geq \cdots \geq v_{i_n}$. Then, $1 - \lambda_2$ is bounded between the graph conductance and the minimal conductance of the rows generating the components of $\boldsymbol{v}$ in sorted order:

$$2 \min_{S \subseteq [n]} \phi(S) \geq 1 - \lambda_2 \geq \frac{1}{2} \left(\min_{1 \leq k \leq n} \phi(\{i_1, i_2, \ldots, i_k\}) \right)^2 . \tag{3}$$

This theorem implies that we can find a good conductance by only searching over the rows generating the components of $\boldsymbol{v}$ in sorted order, rather than over the exponentially many subsets of vertices in the graph. The resultant conductance ϕ has $\phi \leq \sqrt{2}(1 - \lambda_2) \leq 2\sqrt{\mathrm{OPT}}$.

2 Proof of the Spectral Conductance Theorem

Recall $\mathbf{B}$ is non-symmetric, which makes it difficult to analyze the eigenvalues. However, we will show there exists a spectrally equivalent symmetric matrix.

Claim 2.1

Suppose $\mathbf{D}^2 = \mathrm{Diag}(\boldsymbol{\pi})$, then $\mathbf{Q} = \mathbf{D}\mathbf{B}\mathbf{D}^{-1}$ is symmetric and spectrally equivalent to $\mathbf{B}$.

Proof: Since $\mathbf{B}$ is reversible, $\mathbf{D}^2\mathbf{B} = \mathbf{B}^\top\mathbf{D}^2$. Thus, $\mathbf{D}\mathbf{B}\mathbf{D}^{-1} = \mathbf{D}^{-1}\mathbf{B}^\top\mathbf{D}$, so $\mathbf{Q}$ is symmetric.

Suppose $\mathbf{B}\boldsymbol{v} = \lambda\boldsymbol{v}$. Then $\mathbf{D}\mathbf{B}\mathbf{D}^{-1}(\mathbf{D}\boldsymbol{v}) = \lambda(\mathbf{D}\boldsymbol{v})$, so all right eigenvectors of $\mathbf{B}$ can be eigenvectors of $\mathbf{Q}$ by linear transformation. ■

For $\lambda_1 = 1$ we have

$$(\boldsymbol{\pi}^\top\mathbf{D}^{-1})\mathbf{Q} = \boldsymbol{\pi}^\top\mathbf{B}\mathbf{D}^{-1} = \boldsymbol{\pi}\mathbf{D}^{-1}, \tag{4}$$

so $\boldsymbol{\pi}\mathbf{D}^{-1}$ is a left eigenvector of $\mathbf{Q}$. By definition of left eigenvector,

$$\lambda_2 = \max_{\boldsymbol{x}:\boldsymbol{\pi}^\top\mathbf{D}^{-1}\boldsymbol{x}=0} \frac{\boldsymbol{x}^\top\mathbf{D}\mathbf{B}\mathbf{D}^{-1}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}. \tag{5}$$

Let $\boldsymbol{y} = \mathbf{D}^{-1}\boldsymbol{x}$, then

$$\begin{aligned} 1 - \lambda_2 &= \min_{\boldsymbol{x}:\boldsymbol{\pi}^\top\mathbf{D}^{-1}\boldsymbol{x}=0} \frac{\boldsymbol{x}^\top(\mathbf{I} - \mathbf{D}\mathbf{B}\mathbf{D}^{-1})\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} && (6)\\ &= \min_{\boldsymbol{x}:\boldsymbol{\pi}^\top\mathbf{D}^{-1}\boldsymbol{x}=0} \frac{\boldsymbol{x}^\top\mathbf{D}(\mathbf{I} - \mathbf{B})\mathbf{D}^{-1}\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} && (7)\\ &= \min_{\boldsymbol{y}:\boldsymbol{\pi}^\top\boldsymbol{y}=0} \frac{\boldsymbol{y}^\top\mathbf{D}^2(\mathbf{I} - \mathbf{B})\boldsymbol{y}}{\boldsymbol{y}^\top\mathbf{D}^2\boldsymbol{y}}. && (8) \end{aligned}$$

Let us analyze the numerator. Since the row sum of $\mathbf{B}$ is 1,

$$\boldsymbol{y}^\top \mathbf{D}^2(\mathbf{I}-\mathbf{B})\boldsymbol{y} = -\sum_{i\neq j}\pi_i b_{ij} y_i y_j + \sum_i \pi_i(1-b_{ii})y_i^2 \tag{9}$$

$$= -\sum_{i\neq j}\pi_i b_{ij} y_i y_j + \sum_{i\neq j}\pi_i b_{ij}\frac{y_i^2+y_j^2}{2} \tag{10}$$

$$= \sum_{i<j}\pi_i b_{ij}(y_i-y_j)^2. \tag{11}$$

Call the final term $\mathcal{E}(\boldsymbol{y},\boldsymbol{y})$; it is a weighted similarity measure between the $\boldsymbol{y}$'s.

Now, we need to come up with a $\boldsymbol{w}$ to represent the minimum conductance cut S and obtain the bound (in a proof by witness). Let

$$w_i = \begin{cases} \sqrt{\frac{\pi(\overline{S})}{\pi(S)}} & i\in S, \\ -\sqrt{\frac{\pi(S)}{\pi(\overline{S})}} & i\in \overline{S}. \end{cases} \tag{12}$$

Clearly $\boldsymbol{\pi}^\top \boldsymbol{w} = 0$, so it is a candidate for the optimization. Then,

$$1-\lambda_2 \leq \frac{\mathcal{E}(\boldsymbol{w},\boldsymbol{w})}{\sum_i \pi_i w_i^2} \tag{13}$$

$$= \frac{\sum_{i\in S, j\in\overline{S}}\pi_i b_{ij}\left(\sqrt{\frac{\pi(\overline{S})}{\pi(S)}}+\sqrt{\frac{\pi(S)}{\pi(\overline{S})}}\right)^2}{\sum_{i\in S}\pi_i + \sum_{i\in\overline{S}}\pi_i} \tag{14}$$

$$q = \frac{\sum_{i\in S, j\in\overline{S}}\pi_i b_{ij}}{\pi(S)\pi(\overline{S})} \tag{15}$$

$$\leq 2\phi(S), \tag{16}$$

because one of $\pi(S)$ and $\pi(\overline{S})$ must be greater than 1/2.